

# Channel Combining and Splitting for Cutoff Rate Improvement

Erdal Arkan, *Senior Member, IEEE*

**Abstract**—The cutoff rate  $R_0(W)$  of a discrete memoryless channel (DMC)  $W$  is often used as a figure of merit alongside the channel capacity  $C(W)$ . If a channel  $W$  is split into two possibly correlated subchannels  $W_1, W_2$ , the capacity function always satisfies  $C(W_1) + C(W_2) \leq C(W)$ , while there are examples for which  $R_0(W_1) + R_0(W_2) > R_0(W)$ . The fact that cutoff rate can be “created” by channel splitting was noticed by Massey in his study of an optical modulation system. This paper gives a general framework for achieving similar gains in the cutoff rate of arbitrary DMCs by methods of channel combining and splitting. The emphasis is on simple schemes that can be implemented in practice. We give several examples that achieve significant gains in cutoff rate at little extra system complexity. Theoretically, as the complexity grows without bound, the proposed framework is capable of boosting the cutoff rate of a channel to arbitrarily close to its capacity in a sense made precise in the paper. Apart from Massey’s work, the methods studied here have elements in common with Forney’s concatenated coding idea, a method by Pinsker for cutoff rate improvement, and certain coded-modulation techniques, namely, Ungerboeck’s set-partitioning idea and Imai–Hirakawa multilevel coding; these connections are discussed in the paper.

**Index Terms**—Channel combining, channel splitting, coded modulation, concatenated coding, cutoff rate, error exponent, multilevel coding, random-coding exponent, reliability exponent, set partitioning, successive cancellation decoding.

## I. INTRODUCTION

THE *cutoff rate* function for any pair of discrete random variables  $(X, Y)$  with a joint distribution  $P(x, y)$  (denoted  $(X, Y) \sim P(x, y)$  in the sequel) is defined as

$$R_0(X; Y) = -\log \sum_y \left[ \sum_x P(x) \sqrt{P(y|x)} \right]^2$$

and for any three random variables  $(X, Z, Y) \sim P(x, z, y)$ , as

$$R_0(X; Y|Z) = -\log \sum_{z,y} P(z) \left[ \sum_x P(x|z) \sqrt{P(y|x,z)} \right]^2.$$

(All logarithms are to the base 2 throughout.) The function  $R_0(X; Y|Z)$  shares many properties of the mutual information

$$I(X, Y|Z) = \sum_{x,z,y} P(x, z, y) \log \frac{P(y|x, z)}{P(y|z)},$$

listed as follows.

Manuscript received April 27, 2004; revised October 20, 2005. The material in this paper was presented at the 2005 IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

The author is with the Electrical–Electronics Engineering Department, Bilkent University, Ankara, TR-06800 Turkey (e-mail: arkan@ee.bilkent.edu.tr).

Communicate by A. E. Ashikhmin, Associate Editor for Coding Theory.  
Digital Object Identifier 10.1109/TIT.2005.862081

- 1) Nonnegativity.  $R_0(X; Y|Z) \geq 0$  with equality if and only if (iff)  $X$  and  $Y$  are conditionally independent given  $Z$ , i.e.,

$$(X, Y, Z) \sim P_Z(z) P_{X|Z}(x|z) P_{Y|Z}(y|z).$$

- 2) Monotonicity in the first ensemble.  $R_0(XZ; Y) \geq R_0(X, Y)$  with equality iff  $Z$  and  $Y$  are conditionally independent given  $X$ , i.e.,

$$(X, Y, Z) \sim P_X(x) P_{Y|X}(y|x) P_{Z|X}(z|x).$$

- 3) Monotonicity in the second ensemble.  $R_0(X; YZ) \geq R_0(X; Y)$  with equality iff  $X$  and  $Z$  are conditionally independent given  $Y$ , i.e.,

$$(X, Y, Z) \sim P_Y(y) P_{X|Y}(x|y) P_{Z|Y}(z|y).$$

- 4) If

$$(X, Y, Z) \sim P_X(x) P_Z(z) P_{Y|XZ}(y|xz)$$

i.e.,  $X$  and  $Z$  are independent, then  $R_0(X; Y|Z) = R_0(X; YZ)$ .

Mutual information has two other important properties: symmetry,  $I(X; Y|Z) = I(Y; X|Z)$ , and chain rule (additivity),  $I(XZ; Y) = I(X; Y) + I(Z; Y|X)$ . For the  $R_0$  function, neither holds in general. Lack of additivity means that for either inequality

$$R_0(XZ; Y) \geq R_0(X; Y) + R_0(Z; Y|X) \quad (1)$$

there exists an example satisfying that inequality strictly. We will give such examples in the next subsection. The main motivation for this paper is to demonstrate that the possibility of

$$R_0(X; Y) + R_0(Z; Y|X) > R_0(XZ; Y)$$

can be used to gain significant advantages in coding and modulation.

For the most part, we will consider discrete memoryless channels. We write  $W : \mathcal{X} \rightarrow \mathcal{Y}$  to denote a discrete memoryless channel (DMC) with input alphabet  $\mathcal{X}$ , output alphabet  $\mathcal{Y}$ , and transition probability  $W(y|x)$  that output  $y \in \mathcal{Y}$  is received given that input  $x \in \mathcal{X}$  is sent. For such a DMC  $W$ , let  $Q$  be a probability distribution on  $\mathcal{X}$ , and let  $(X, Y) \sim Q(x)W(y|x)$ . The cutoff rate of  $W$  under input distribution  $Q$  is defined as  $R_0(Q; W) = R_0(X; Y)$ . The cutoff rate of  $W$  is defined as

$$R_0(W) = \max_Q R_0(Q; W).$$

### A. Lack of Chain Rule for the Cutoff Rate

We give two examples showing that no general inequality exists between the left- and right-hand sides of (1). The first example is derived from Massey’s work [1] and illustrates in a simple fashion the main goals of the present paper. In a study of

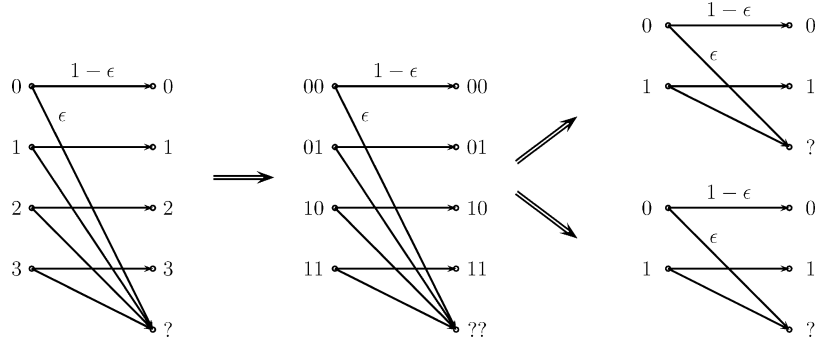


Fig. 1. Relabeling and splitting of a QEC into two correlated BECs.

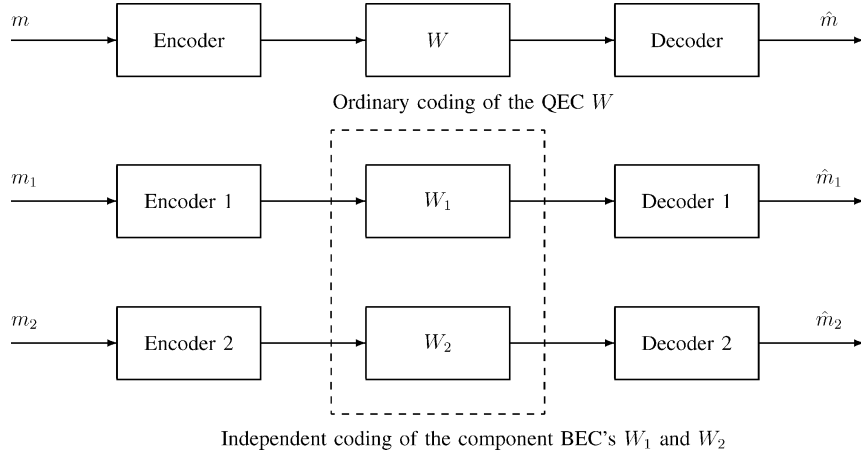


Fig. 2. Two coding alternatives over the QEC.

coding and modulation for an optical communication system, Massey showed that splitting a given channel into correlated subchannels may lead to an improvement in the sum cutoff rate. He modeled the optical communication channel as an  $M$ -ary erasure channel and considered splitting the  $M$ -ary channel into binary erasure channels (BECs). The following example shows Massey's idea for  $M = 4$ . This same example was also discussed in [2] to illustrate some unexpected behavior of cutoff rates.

*Example 1 (Massey [1]):* Consider the quaternary erasure channel (QEC), shown on the left in Fig. 1, and relabel its inputs and outputs to obtain the channel  $W : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}_1 \times \mathcal{Y}_2$  where  $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ ,  $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, ?\}$ , and

$$W(y_1 y_2 | x_1 x_2) = \begin{cases} 1 - \epsilon, & y_1 y_2 = x_1 x_2 \\ \epsilon, & y_1 y_2 = ?? \end{cases}$$

where  $0 \leq \epsilon \leq 1$  is the erasure probability. The QEC  $W$  can be decomposed into two BECs:  $W_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i$ ,  $i = 1, 2$ , as shown on the right in Fig. 1. In this decomposition, a transition  $(x_1, x_2) \rightarrow (y_1, y_2)$  over the QEC is viewed as two transitions,  $x_1 \rightarrow y_1$  and  $x_2 \rightarrow y_2$ , taking place on the respective component channels, with

$$W_i(y_i | x_i) = \begin{cases} 1 - \epsilon, & y_i = x_i \\ \epsilon, & y_i = ? \end{cases}$$

These BECs are fully correlated in the sense that an erasure occurs either in both or in none.

We now compare the two coding alternatives for the QEC that are shown in Fig. 2 in terms of their respective capacities and

cutoff rates. The first alternative is the ordinary coding of the QEC as a single-user channel. The second is the independent coding of the component BECs  $W_1$  and  $W_2$ , ignoring the correlation between the two subchannels. The capacity and cutoff rate of an  $M$ -ary erasure channel are given by  $C = (1 - \epsilon) \log M$  and  $R_0 = \log M - \log[1 + (M - 1)\epsilon]$ , respectively, for any  $M \geq 1$ . It follows that  $C(W) = C(W_1) + C(W_2)$ , i.e., the capacity of the QEC is not degraded by splitting it into two BECs. On the other hand, it is easily verified and shown in Fig. 3 that the cutoff rate is improved by splitting, i.e.,  $R_0(W_1) + R_0(W_2) > R_0(W)$ .

Massey's example shows that the left-hand side in (1) may be strictly smaller than the right-hand side. To see this, define an ensemble

$$(X_1, X_2, Y_1, Y_2) \sim Q(x_1)Q(x_2)W(y_1 y_2 | x_1 x_2)$$

where  $Q$  is the uniform distribution on  $\{0, 1\}$ , and verify that

$$R_0(W) = R_0(X_1 X_2; Y_1 Y_2)$$

$$R_0(W_1) = R_0(X_1; Y_1) = R_0(X_1; Y_1 Y_2)$$

and

$$R_0(W_2) = R_0(X_2; Y_2) = R_0(X_2; Y_1 Y_2 | X_1).$$

Implications of the possibility that cutoff rate can be "created" will be discussed shortly. We complete this subsection by giving an example for the reverse inequality in (1).

*Example 2:* Let  $W : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}_1 \times \mathcal{Y}_2$  where  $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ ,  $\mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1, ?\}$ , and

$$W(y_1 y_2 | x_1 x_2) = \begin{cases} 1/2, & y_1 y_2 = x_1 ? \text{ or } y_1 y_2 = ? x_2 \\ 0, & \text{otherwise.} \end{cases}$$

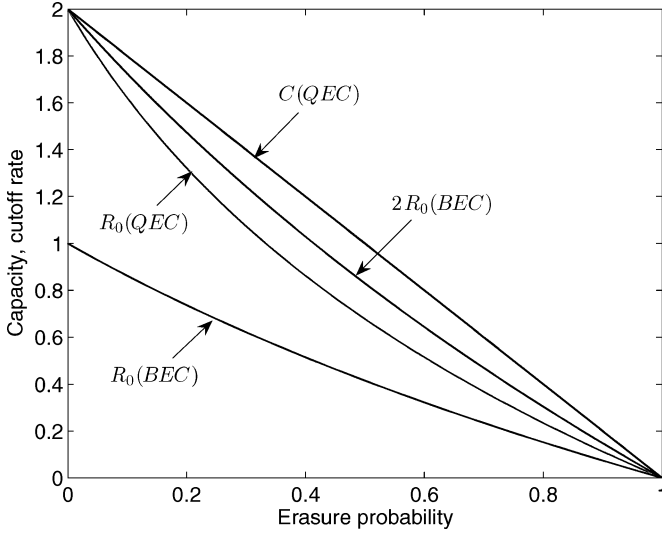


Fig. 3. Capacity and cutoff rate for the splitting of a QEC.

The channel  $W$  may be thought of as consisting of two BECs  $W_1 : \mathcal{X}_1 \rightarrow \mathcal{Y}_1$ ,  $W_2 : \mathcal{X}_2 \rightarrow \mathcal{Y}_2$ , where each has erasure probability  $\epsilon = 1/2$ . Subchannels  $W_1$  and  $W_2$  are correlated since an erasure occurs either in one or the other but never in both.

$R_0(W)$  is achieved by the uniform distribution  $Q$  on  $\mathcal{X}_1 \times \mathcal{X}_2$  and  $R_0(W) = 1$ . On the other hand, we have

$$R_0(W_1) = R_0(W_2) = \log(4/3).$$

Thus, this is an example where the sum cutoff rate is worsened by splitting,  $R_0(W_1) + R_0(W_2) < R_0(W)$ . This example also shows that the left-hand side in (1) may be strictly greater than the right-hand side. This can be seen by considering the ensemble  $(X_1, X_2, Y_1, Y_2) \sim Q(x_1)Q(x_2)W(y_1 y_2 | x_1 x_2)$  where  $Q$  is the uniform distribution on  $\{0, 1\}$ , and verifying that

$$R_0(W) = R_0(X_1 X_2; Y_1 Y_2)$$

$$R_0(W_1) = R_0(X_1; Y_1) = R_0(X_1; Y_1 Y_2)$$

and

$$R_0(W_2) = R_0(X_2; Y_2) = R_0(X_2; Y_1 Y_2 | X_1).$$

In both of the above examples, the erasure events in the subchannels  $W_1 : \mathcal{X}_1 \rightarrow \mathcal{Y}_1$  and  $W_2 : \mathcal{X}_2 \rightarrow \mathcal{Y}_2$  are fully correlated; the correlation is positive in the first, negative in the second. This vaguely suggests that, in order to obtain cutoff rate gains, one should seek to split a given channel into subchannels in such a way that the noise levels in subchannels are positively correlated.

### B. Significance of the Cutoff Rate

Channel cutoff rate has long been used as a figure of merit for coding and modulation systems and this has been justified in well-known works, such as [3] and [4]. Here, we summarize the argument in favor of using  $R_0$  as a figure of merit.

One reason for the significance of  $R_0$  stems from its role in connection with *sequential decoding*, which is a decoding algorithm for tree codes invented by Wozencraft [5], with important later contributions by Fano [6]. Sequential decoding can be used

to achieve arbitrarily reliable communication on any DMC  $W$  at rates arbitrarily close to the cutoff rate  $R_0(W)$  while keeping the average computation per decoded digit bounded by a constant that depends on the code rate, the channel  $W$ , but not on the desired level of reliability. If the desired rate is between  $R_0(W)$  and channel capacity  $C(W)$ , sequential decoding can still achieve arbitrarily reliable communication but the average computation becomes arbitrarily large as well. Proofs of these results can be found in the textbooks [7]–[9].

If we consider using sequential decoders in the two coding alternatives in Fig. 2, the achievable sum cutoff rate by the second alternative exceeds that by the first, since

$$2R_0(\text{BEC}) > R_0(\text{QEC}), \quad \text{for all } 0 < \epsilon < 1.$$

This improvement comes at negligible extra system complexity, if any.

Apart from its significance in sequential decoding,  $R_0$  appears in the union bound on the probability of error for coding and modulation systems [7, p. 396], and hence, serves as a simple measure of reliability. This issue is connected to the relation of  $R_0$  to the channel reliability exponent, and will be discussed in detail in Section VII.

### C. Outline

This paper addresses the following questions raised by Massey's example. Can *any* DMC be split in some way to achieve coding gains as measured by improvements in the cutoff rate? And, if so, what are the limits of such gains?

We address these questions in a framework where channel combining prior to splitting is allowed. In Massey's example there is no channel combining; a given channel is simply split into two subchannels. In general, however, it may not be possible to split a given channel into subchannels so as to obtain a gain in the cutoff rate. By applying channel combining prior to splitting, we manufacture large channels which can be more readily split to yield cutoff rate gains. In fact, we show that the cutoff rate of a given DMC  $V$  can be improved all the way to its capacity  $C(V)$  in the limit of combining an arbitrarily large number of independent copies of  $V$ .

An outline of the rest of the paper follows. In Section II, we describe the general framework for cutoff rate improvement. Section III demonstrates the effectiveness of the proposed framework by giving two very simple examples in which the cutoff rates of BECs and BSCs are improved significantly. Section IV explores the limits of practically achievable cutoff rate gains by combining a moderately large number of BSCs based on some *ad hoc* channel combining methods. Section V exhibits Pinsker's method [10] as a special case of the method presented in Section II, thus establishing that the cutoff rate of any DMC can be boosted to as close to channel capacity as desired. (Unfortunately, this is an asymptotic result with no immediate practical significance.) In Section VI, we give an example that illustrates the common elements of the channel splitting method and coded-modulation schemes. In Section VII, we show that channel splitting may be used to improve the reliability–complexity tradeoff in maximum-likelihood (ML) decoding. Section VIII concludes the paper with a summary and discussion.

The present work is tied to previous work along many threads. We have already mentioned Massey's work [1] as the main starting point. There is also a close similarity to techniques employed in coded modulation and multilevel coding as represented by the pioneering works of Ungerboeck [11], [12] and Imai–Hirakawa [13]. Finally, the present work is connected through Pinsker's serial concatenation approach [10] to iterative coding methods that date back to Elias [14]. Concatenated coding in general is a well-known technique for improving the reliability–complexity tradeoff in coding, and it is pertinent to cite Forney's seminal work on the subject [15] and important later works [16], [17] as part of the broader background for the present paper.

*Notation:* We write  $a_i^j$  to denote the vector  $(a_i, \dots, a_j)$  for any  $1 \leq i \leq j$ . If  $j < i$ ,  $a_i^j$  denotes a void vector.

## II. CHANNEL COMBINING AND SPLITTING

In order to seek coding gains in the spirit of Massey's example, we will consider DMCs of the form  $W : \mathcal{X}^n \rightarrow \mathcal{Y}^n$  for some integer  $n \geq 2$ . Such a channel may have been obtained by labeling the input and output symbols of a given channel by vectors, as in Massey's example. Typically, such a channel will be obtained by combining  $n$  independent copies of a given DMC  $V : \mathcal{X} \rightarrow \mathcal{Y}$ , as shown in Fig. 4. We incorporate into the channel combining procedure a bijective *label mapping* function  $f : \mathcal{X}^n \rightarrow \mathcal{X}^n$ . Relabeling of the inputs of  $V^{\otimes n}$  (the channel that consists of  $n$  independent uses of  $V$ ) is an essential ingredient of the proposed method; without such a relabeling, there would be no gain. The resulting channel is a DMC  $W : \mathcal{X}^n \rightarrow \mathcal{Y}^n$  such that

$$W(y_1^n | u_1^n) = \prod_{i=1}^n V(y_i | x_i)$$

where  $x_1^n = f(u_1^n)$ .

Whatever the origin of the channel  $W : \mathcal{X}^n \rightarrow \mathcal{Y}^n$ , we regard  $W$  as a channel with  $n$  input terminals where each terminal is encoded by a different user. We will consider splitting  $W$  into  $n$  subchannels by using the coding system shown in Fig. 5 where a type of decoder known variously as a *successive cancellation* decoder or a *multilevel* decoder is employed. The latter term is more common in the coded-modulation literature [13]. We will adopt the former term in the sequel since the context here is broader than modulation.

The decoder in a successive cancellation system is designed around a random code ensemble for channel  $W$ , specified by a random vector  $U_1^n \sim Q_1(u_1) \cdots Q_n(u_n)$ , where  $Q_i$  is a probability distribution on  $\mathcal{X}$ ,  $1 \leq i \leq n$ . Roughly,  $U_i$  corresponds to the input random variable that is transmitted at the  $i$ th input terminal. Given such a reference ensemble, we define  $n$  DMCs  $W_i : \mathcal{X} \rightarrow \mathcal{Y}^n \times \mathcal{X}^{i-1}$ ,  $1 \leq i \leq n$ , so that

$$W_i(y_1^n, u_1^{i-1} | u_i) = Q_1(u_1) \cdots Q_{i-1}(u_{i-1}) \times \sum_{u_{i+1}^n} Q_{i+1}(u_{i+1}) \cdots Q_n(u_n) W(y_1^n | u_1^n).$$

The  $i$ th decoder in the system will use  $W_i$  as its channel model in carrying out the decoding task.

The precise operation of the successive cancellation system is as follows. User  $i$  encodes a message  $m_i$  into a vector

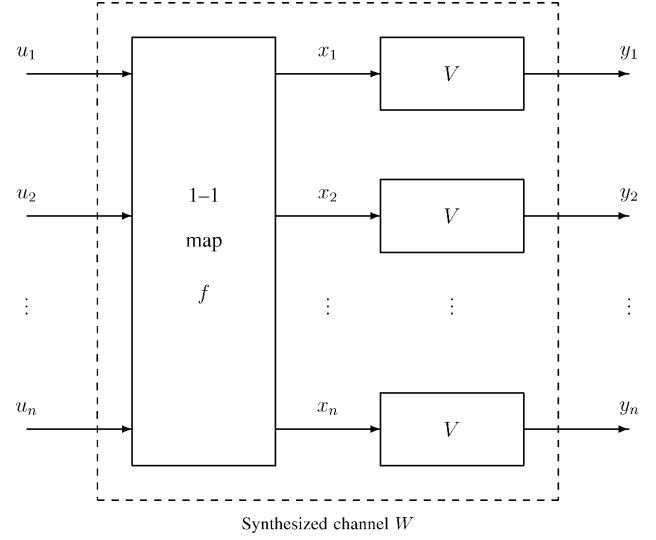


Fig. 4. Channel combining and input relabeling.

$\mathbf{u}_i = (u_{i,1}, \dots, u_{i,T}) \in \mathcal{X}^T$ , which it transmits at input terminal  $i$ ,  $1 \leq i \leq n$ . The codeword length  $T$  is assumed to be the same for all users to simplify the description; in general, each user may have a different codeword length. The messages  $m_1, \dots, m_n$  are assumed independent. Decoder 1 observes the channel output vector  $\mathbf{y} = (y_{i,t} : i = 1, \dots, n; t = 1, \dots, T)$  where  $y_{i,t}$  is the output of the  $i$ th copy of  $V$  at time  $t$ , and generates its estimate  $\hat{m}_1$  of  $m_1$ , and also an estimate  $\hat{\mathbf{u}}_1 = (\hat{u}_{1,1}, \dots, \hat{u}_{1,T})$  of  $\mathbf{u}_1$ . It puts out  $\hat{m}_1$  as its decision, and feeds  $\hat{\mathbf{u}}_1$  to all succeeding decoders to help them with their decoding tasks. This first decoder knows only the code used by encoder 1 and uses the channel model  $W_1$ , in effect, modeling the inputs  $u_{2,t}, \dots, u_{n,t}$  at time  $t$  at other input terminals as generated by a memoryless source from the distribution  $Q_2(u_{2,t}) \cdots Q_n(u_{n,t})$ ,  $1 \leq t \leq T$ . (In case the users employ tree coding with certain truncation periods, each user is assumed to be aware of the position and identity of the truncation symbols used by each succeeding user and to modify the channel model accordingly.) In general, decoder  $i$  observes  $(\mathbf{y}, \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_{i-1})$  and proceeds to generate an estimate  $\hat{m}_i$  of  $m_i$  and an estimate  $\hat{\mathbf{u}}_i = (\hat{u}_{i,1}, \dots, \hat{u}_{i,T})$  of  $\mathbf{u}_i$  using the channel model  $W_i(y_{1,t}, \dots, y_{n,t}, \hat{u}_{1,t}, \dots, \hat{u}_{i-1,t} | u_{i,t})$ . The estimate  $\hat{m}_i$  is sent out, while  $\hat{\mathbf{u}}_i$  is passed to the succeeding decoders.

This successive cancellation system provides the general framework in which we will consider channel combining and splitting methods. We will assume that the decoders in the system are sequential decoders and use the sum cutoff rate as the criterion for measuring coding gains. By standard random-coding arguments for tree codes and sequential decoding, one may show that an ensemble average error probability

$$\bar{P}_e = \Pr[(\hat{m}_1, \dots, \hat{m}_n) \neq (m_1, \dots, m_n)]$$

as small as desired can be achieved while still ensuring that the average decoding complexity per decoded symbol is bounded by a constant provided that encoder  $i$  operates at a rate  $R_i < R_0(Q_i; W_i)$ ,  $1 \leq i \leq n$ . Thus, the sum rate achievable by such a scheme is given by  $\sum_{i=1}^n R_0(Q_i; W_i)$ .

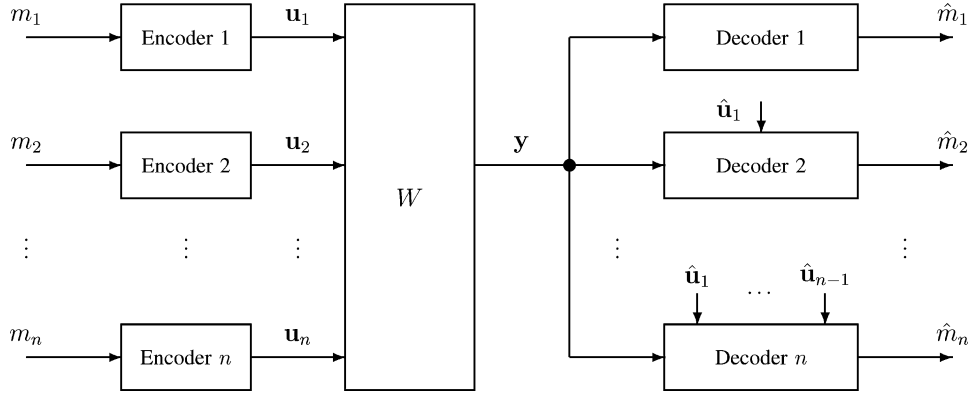


Fig. 5. Channel splitting by successive cancellation.

Note that  $R_0(Q_i; W_i)$  equals  $R_0(U_i; Y_1^n U_1^{i-1})$ , which in turn equals  $R_0(U_i; Y_1^n | U_1^{i-1})$  since  $U_i$  and  $U_1^{i-1}$  are independent. So, the sum cutoff rate for the preceding scheme normalized by the number of channels is

$$\bar{R}_0(U_1^n; Y_1^n) \triangleq \frac{1}{n} \sum_{i=1}^n R_0(U_i; Y_1^n | U_1^{i-1}).$$

We will speak of a coding gain if  $\bar{R}_0(U_1^n; Y_1^n)$  is greater than  $R_0(V)$ , the ordinary cutoff rate of the elementary channel  $V$ .

It may be of interest to consider also the cutoff rate for the channel  $W$ , namely

$$R_0(W) = \max_Q R_0(Q; W)$$

where the maximum is over all  $Q(u_1^n)$ , not necessarily in product form. The rate  $R_0(W)$  is the supremum of achievable rates by a *single* sequential decoder that decodes all inputs  $u_1^n$  jointly, with no restriction that different inputs of  $W$  are encoded independently. Gallager's "parallel channels theorem" [8, p. 149] ensures that  $R_0(W) \leq nR_0(V)$  with equality iff the label map  $f$  is bijective. Thus, the only possible method of achieving a coding gain as measured by the sum cutoff rate is to split the decoding function into *multiple* sequential decoders. The splitting of the encoder plays an incidental role in improving the sum cutoff rate by facilitating the use of multiple sequential decoders in a successive cancellation configuration.

### III. BEC AND BSC EXAMPLES

In this section, we give an immediate application of the method outlined in the preceding section to the BEC and BSC. The main point of this section is to illustrate that by combining just two copies of these channels one can obtain significant improvements in the cutoff rate.

*Example 3 (BEC):* Let  $V : \mathcal{X} \rightarrow \mathcal{Y}$  be the BEC with alphabets  $\mathcal{X} = \{0, 1\}$ ,  $\mathcal{Y} = \{0, 1, ?\}$ , and erasure probability  $\epsilon$ . Consider combining two copies of  $V$  to obtain a channel  $W : \mathcal{X}^2 \rightarrow \mathcal{Y}^2$  as shown in Fig. 6. The label map is given by

$$f : (u_1, u_2) \rightarrow (x_1, x_2) = (u_1 \oplus u_2, u_2)$$

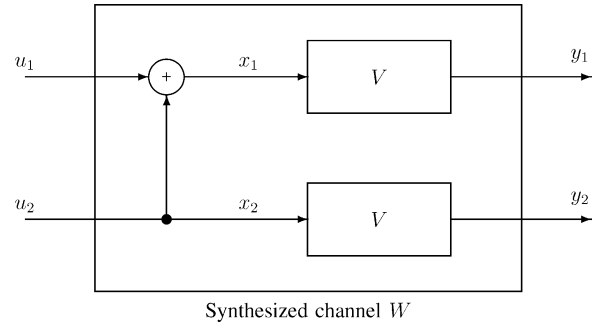


Fig. 6. Synthesis of a quaternary input channel from two binary-input channels.

where  $\oplus$  denotes modulo-2 addition. Let the input variables be specified as  $U_1^2 \sim Q_1(u_1)Q_2(u_2)$  where  $Q_1, Q_2$  are uniform on  $\{0, 1\}$ . Then

$$\begin{aligned} R_0(U_1; Y_1^2) &= 1 - \log(1 + 2\epsilon - \epsilon^2) \\ R_0(U_2; Y_1^2 | U_1) &= 1 - \log(1 + \epsilon^2). \end{aligned}$$

A heuristic interpretation of these cutoff rates can be given by observing that user 1's channel  $u_1 \rightarrow (y_1, y_2)$  is effectively a BEC with erasure probability  $1 - (1 - \epsilon)^2 = 2\epsilon - \epsilon^2$ ; an erasure occurs in this channel when either  $x_1$  or  $x_2$  is erased. On the other hand, given that decoder 2 is supplied with the correct value of  $u_1$ , the channel seen by user 2 is a BEC with erasure probability  $\epsilon^2$ ; an erasure occurs only when both  $x_1$  and  $x_2$  are erased. The normalized sum cutoff rate under this scheme is given by

$$\bar{R}_0(U_1^2; Y_1^2) = 1 - \frac{1}{2} [\log(1 + 2\epsilon - \epsilon^2) + \log(1 + \epsilon^2)]$$

which is to be compared with the ordinary cutoff rate of the BEC,  $R_0(V) = 1 - \log(1 + \epsilon)$ . These cutoff rates are shown in Fig. 7. The figure shows and it can be verified analytically that the above method improves the cutoff rate for all  $0 < \epsilon < 1$ .

We note that the distribution of  $U_1^n$  and the labeling  $f$  given above are optimal in the sense that it maximizes  $\bar{R}_0(U_1^n; Y_1^n)$ . This can be verified by exhaustively trying all  $4^n$  possibilities for  $f$ . In order to achieve higher coding gains, one needs to combine a larger number of copies of the BEC.

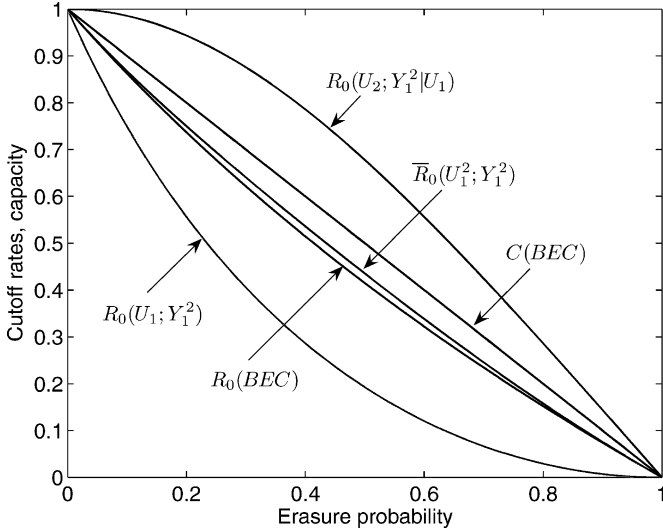


Fig. 7. Cutoff rates for the splitting of BEC.

*Example 4 (BSC):* Let  $V : \mathcal{X} \rightarrow \mathcal{Y}$  be a BSC with  $\mathcal{X} = \mathcal{Y} = \{0, 1\}$  and

$$V(y|x) = \begin{cases} 1 - \epsilon, & y = x \\ \epsilon, & y \neq x \end{cases}$$

for all  $x \in \mathcal{X}, y \in \mathcal{Y}$ . Assume  $0 \leq \epsilon \leq 1/2$ . The cutoff rate of the BSC is given by

$$R_0(V) = 1 - \log(1 + \gamma(\epsilon))$$

where  $\gamma(\delta) \triangleq \sqrt{4\delta(1-\delta)}$  for  $0 \leq \delta \leq 1$ .

Combine two copies of the BSC as in Fig. 6 to obtain the channel  $W$  and define the input variables  $U_1^2 \sim Q_1(x_1)Q_2(x_2)$  where  $Q_1, Q_2$  are uniform on  $\{0, 1\}$ . The cutoff rates  $R_0(U_1; Y_1^2)$  and  $R_0(U_2; Y_1^2|U_1)$  can be obtained by direct calculation; however, it is instructive to obtain them by the following argument. The input and output variables of the channel  $W$  are related by

$$y_1 = u_1 \oplus u_2 \oplus e_1$$

$$y_2 = u_2 \oplus e_2$$

where  $e_1$  and  $e_2$  are independent noise terms taking the values 0 and 1 with probabilities  $1 - \epsilon$  and  $\epsilon$ , respectively. Notice that  $y_1 \oplus y_2$  is a sufficient statistic for decoder 1, whose goal is to estimate  $u_1$ . Thus, the channel that decoder 1 sees is, in effect,  $u_1 \rightarrow u_1 \oplus e_1 \oplus e_2$ , which is a BSC with crossover probability  $\epsilon_2 = 2\epsilon(1 - \epsilon)$  and has cutoff rate

$$R_0(U_1; Y_1^2) = 1 - \log(1 + \gamma(\epsilon_2)).$$

Next, note that the channel  $u_2 \rightarrow (y_1, y_2)$ , given the knowledge of  $u_1$ , is equivalent to the channel

$$u_2 \rightarrow (y_1 \oplus u_1, y_2) = (u_2 \oplus e_1, u_2 \oplus e_2)$$

which is a BSC with diversity order 2 and has cutoff rate

$$R_0(U_2; Y_1^2|U_1) = 1 - \log(1 + \gamma(\epsilon)^2).$$

Thus, the normalized sum cutoff rate with this splitting scheme is given by

$$\bar{R}_0(U_1^2; Y_1^2) = 1 - \frac{1}{2} [\log(1 + \gamma(\epsilon_2)) + \log(1 + \gamma(\epsilon)^2)]$$

which is larger than  $R_0(V)$  for all  $0 < \epsilon < 0.5$ , as shown in Fig. 8.

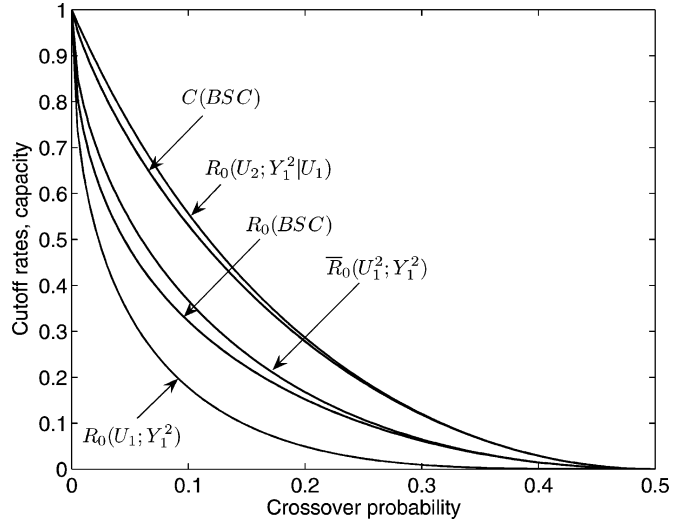


Fig. 8. Cutoff rates for the splitting of BSC.

#### IV. LINEAR LABEL MAPS

In the previous section, we have shown that there exist linear label maps of order  $n = 2$  that improve the sum cutoff rate. In this section, we give higher order linear label maps that attain further improvements. Throughout the section, we restrict attention to a BSC with a crossover probability  $0 \leq \epsilon \leq 1/2$ .

We consider combining  $n$  copies of a BSC as in Fig. 4 using a label map  $f$  that is linear. Thus, we set  $x_1^n = u_1^n F$  where  $F$  is a full-rank matrix of size  $n \times n$ . The output of the combined channel  $W$  has the form  $y_1^n = x_1^n + e_1^n$  where  $e_1^n$  is the channel noise vector with independent and identically distributed (i.i.d.) components.

For capacity and cutoff rate calculations, we use an input ensemble  $U_1^n$  consisting of i.i.d. components, each component equally likely to take the values 0 and 1. This defines a joint ensemble  $U_1^n \rightarrow X_1^n \rightarrow Y_1^n$  where  $X_1^n$  is the channel input and  $Y_1^n$  the channel output.

We consider a successive cancellation decoder as in Fig. 5. Note that each decoder observes the entire channel output vector  $y_1^n$  and can compute the vector  $v_1^n = y_1^n F^{-1} = u_1^n + e_1^n F^{-1}$ . We denote the  $i$ th column of  $F^{-1}$  by  $c_i$ , so we have  $v_i = y_1^n c_i$ ,  $i = 1, \dots, n$ . For decoder 1, which is interested in recovering  $u_1$  only, a sufficient statistic is the first component of  $v_1^n$ , namely

$$v_1 = y_1^n c_1 = u_1 + e_1^n c_1.$$

If the number of 1's in  $c_1$  equals  $k$ , then the channel  $u_1 \rightarrow v_1$  is the cascade of  $k$  independent BSCs, and is itself a BSC with a crossover probability

$$\epsilon_k = \frac{1 - (1 - 2\epsilon)^k}{2}.$$

The cutoff rate of this channel equals  $R_0(U_1; V_1) = R_0(U_1; Y_1^n)$ , and is given in Fig. 9 as a function of  $\epsilon$  for various  $k$ .

Supposing that the correct value of  $u_1$  is supplied to decoder 2, that decoder in turn can compute the left-hand sides of the two equations

$$v_1 \oplus u_1 = e_1^n c_1$$

$$v_2 = u_2 \oplus e_1^n c_2$$

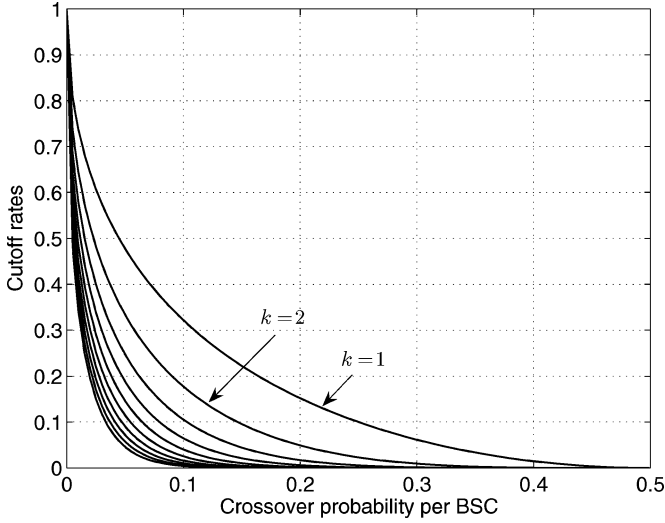


Fig. 9. Cutoff rate for the cascade of  $k = 1, \dots, 10$  BSCs versus the crossover probability per BSC.

which form sufficient statistics for estimating  $u_2$ . This channel has cutoff rate  $R_0(U_2; V_1^2 | U_1) = R_0(U_2; Y_1^n | U_1)$ . In general, supposing that  $u_1^{i-1}$  have been supplied to decoder  $i$ ,  $1 \leq i \leq n$ , decoder  $i$  computes the sufficient statistics

$$\begin{aligned} v_1 \oplus u_1 &= e_1^n c_1 \\ &\vdots \\ v_{i-1} \oplus u_{i-1} &= e_1^n c_{i-1} \\ v_i &= u_i \oplus e_1^n c_i \end{aligned}$$

and estimates  $u_i$ , achieving a cutoff rate

$$R_0(U_i; V_1^i | U_1^{i-1}) = R_0(U_i; Y_1^n | U_1^{i-1}).$$

The following example illustrates the above method.

*Example 5:* Suppose

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

This matrix equals its inverse  $F^{-1} = F$ . Thus, the first decoder has the channel

$$v_1 = y_1 \oplus y_2 \oplus y_3 \oplus y_4 = u_1 \oplus e_1 \oplus e_2 \oplus e_3 \oplus e_4.$$

This is a BSC with crossover probability  $\epsilon_4 = 4\epsilon(1-\epsilon)^3 + 4(1-\epsilon)^3\epsilon$  and has cutoff rate  $R_0(U_1; Y_1^4) = 1 - \log(1 + \gamma(\epsilon_4))$  where  $\gamma(\delta)$  is as defined before. Supposing that  $u_1$  is decoded correctly and passed to the second decoder, decoder 2 has the channel

$$\begin{aligned} v_1 \oplus u_1 &= e_1 \oplus e_2 \oplus e_3 \oplus e_4 \\ v_2 &= u_2 \oplus e_2 \oplus e_4 \end{aligned}$$

or, alternatively, by adding the second equation to the first

$$\begin{aligned} v_1 \oplus v_2 \oplus u_1 &= u_2 \oplus e_1 \oplus e_3 \\ v_2 &= u_2 \oplus e_2 \oplus e_4 \end{aligned}$$

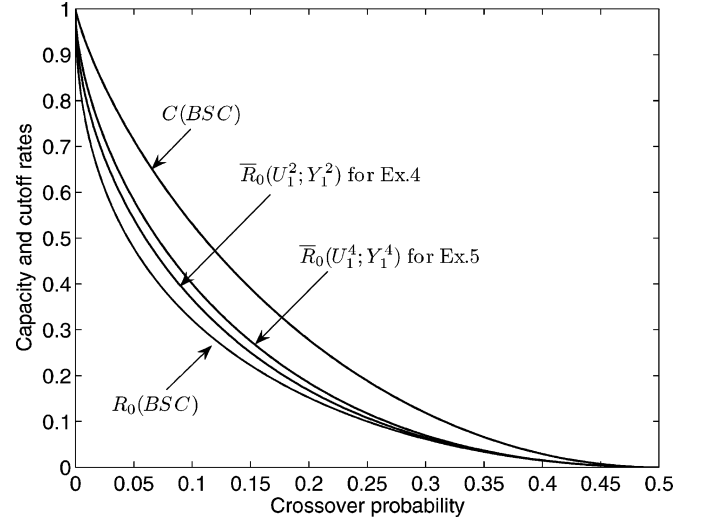


Fig. 10. Capacity and cutoff rates for Example 5.

from which it tries to decode  $u_2$ . This is a channel where  $u_2$  is observed through two independent BSCs each with crossover probability  $\epsilon_2 = 2\epsilon(1-\epsilon)$  and has cutoff rate

$$R_0(U_2; Y_1^4 | U_1) = 1 - \log(1 + \gamma(\epsilon_2)^2).$$

Given that both  $u_1$  and  $u_2$  are available to decoder 3, it has the channel

$$\begin{aligned} v_1 \oplus u_1 &= e_1 \oplus e_2 \oplus e_3 \oplus e_4 \\ v_2 \oplus u_2 &= e_2 \oplus e_4 \\ v_3 &= u_3 \oplus e_3 \oplus e_4. \end{aligned}$$

By adding the third equation to the first and second, we see that this is a channel where the variable  $u_3$  is observed through three channels, each of which is a BSC with crossover probability  $\epsilon_2$ ; however, in this instance, the channels are not independent. The cutoff rate  $R_0(U_3; Y_1^4 | U_1^2)$  can still be computed readily.

Finally, given  $u_1^3$ , decoder 4 has the channel

$$\begin{aligned} v_1 \oplus u_1 &= e_1 \oplus e_2 \oplus e_3 \oplus e_4 \\ v_2 \oplus u_2 &= e_2 \oplus e_4 \\ v_3 \oplus u_3 &= e_3 \oplus e_4 \\ v_4 &= u_4 \oplus e_4 \end{aligned}$$

or, equivalently

$$\begin{aligned} v_1 \oplus v_2 \oplus v_3 \oplus v_4 \oplus u_1 \oplus u_2 \oplus u_3 &= u_4 \oplus e_1 \\ v_2 \oplus v_4 \oplus u_2 &= u_4 \oplus e_2 \\ v_3 \oplus v_4 \oplus u_3 &= u_4 \oplus e_3 \\ v_4 &= u_4 \oplus e_4. \end{aligned}$$

This is a channel where  $u_4$  is observed through four independent BSCs, and

$$R_0(U_4; Y_1^4 | U_1^3) = 1 - \log(1 + \gamma(\epsilon)^4).$$

The normalized sum cutoff rate  $\bar{R}_0(U_1^4, Y_1^4)$  is plotted in Fig. 10 as a function of  $\epsilon$ ; also shown in the figure are the cutoff rate  $R_0$  and capacity  $C$  of the raw BSC. There is a visible improvement in the sum cutoff rate compared to the scheme in

Example 4 and the gap between  $R_0$  and  $C$  is roughly half-way closed for  $\epsilon \approx 0.1$ .

This improvement in cutoff rate is achieved at the expense of somewhat increased system complexity. One element of complexity for a successive cancellation decoder is the size of the transition probability matrices at each level. This matrix is needed to compute the likelihood ratios in ML decoding or the *metric* values in sequential decoding. In this example, the channel at level  $i$ ,  $1 \leq i \leq 4$ , has two inputs and  $2^i$  outputs; the corresponding transition probability matrix has  $2^{i+1}$  entries. In general, the storage complexity of these matrices grows exponentially in the number  $n$  of channels combined, and sets a limit on practical applicability of the method for large  $n$ .

#### A. Kronecker Powers of a Given Labeling

The linear map  $F$  in the above example has the form  $F = A \otimes A$  where

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

is the linear map used in Example 4. We have also carried out cutoff rate calculations for linear maps of the form  $F = A^{\otimes k}$  for  $k = 1, 2, 3, 4$  at a fixed raw error probability  $\epsilon = 0.1$ . The resulting normalized sum cutoff rates are listed in the following table. For comparison, the cutoff rate and capacity of the BSC at  $\epsilon = 0.1$  are  $R_0 = .3219$  and  $C = .5310$ .

$k$	$\bar{R}_0$
1	0.3670
2	0.4016
3	0.4245
4	0.4433

The scheme with  $F = A^{\otimes k}$  has  $n = 2^k$  subchannels. The number of possible values of the output vector  $y_1^n$  equals  $2^{2^k}$ . The rapid growth of this number prevented computing  $\bar{R}_0$  for  $k \geq 5$ .

#### B. Label Maps From Block Codes

Let  $G = [P \ I_k]$  be the generator matrix, in systematic form, of an  $(n, k)$  linear binary block code  $\mathcal{C}$ . Here,  $P$  is a  $k \times (n - k)$  matrix and  $I_k$  is the  $k$ -dimensional identity matrix. A linear label map is obtained by setting

$$F = \left[ \begin{array}{c|c} I_{n-k} & 0 \\ \hline P & I_k \end{array} \right]. \quad (2)$$

Note that  $F$  has full rank and  $F^{-1} = F$ . Also note that the first  $(n - k)$  columns of  $F$  form the transpose  $H^T$  of a *parity-check* matrix  $H = [I_{n-k} \ P^T]$  for  $\mathcal{C}$ . Thus, when the receiver computes the vector

$$v_1^n = y_1^n F^{-1} = y_1^n F$$

the first  $(n - k)$  coordinates of  $v_1^n$  have the form

$$v_i = u_i \oplus s_i, \quad 1 \leq i \leq n - k \quad (3)$$

where  $s_i$  is the  $i$ th element of the *syndrome* vector

$$s_1^k \triangleq y_1^n H^T = e_1^n H^T.$$

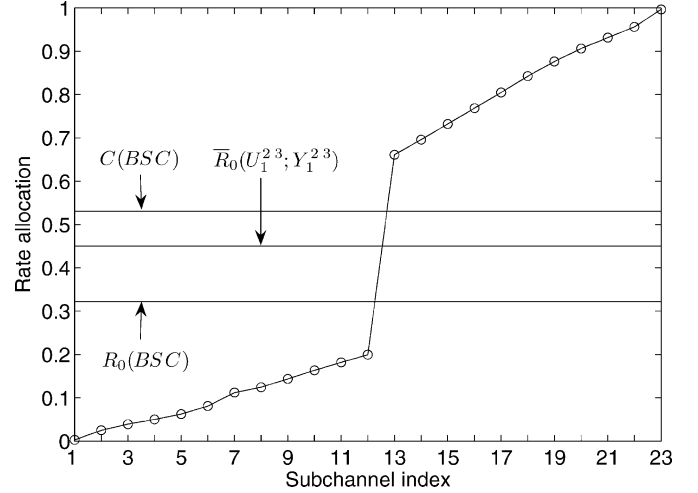


Fig. 11. Rate allocation for Example 6.

The  $i$ th “syndrome subchannel” (3) is a BSC with crossover probability  $\epsilon_k$  where  $k$  is the number of 1’s in the  $i$ th row of  $H$ . The remaining subchannels, which we call “information subchannels,” have the form

$$v_i = u_i \oplus e_i, \quad (n - k + 1) \leq i \leq n.$$

The motivation for the above method is to make use of linear block codes with good distance properties to improve the sum cutoff rate. The following example illustrates the method.

*Example 6 (Dual of the Golay Code):* Let  $F$  be as in (2) with  $n = 23$ ,  $k = 11$ , and

$$P = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The code with the generator matrix  $G = [P \ I_{11}]$  is the dual of the Golay code [18], [19, p. 119]. We computed the normalized sum cutoff rate  $\bar{R}_0(U_1^{23}; Y_1^{23}) = 0.4503$  at  $\epsilon = 0.1$  for this scheme. The rate allocation vector

$$(R_0(U_i; Y_1^{23} | U_1^{i-1}) : 1 \leq i \leq 23)$$

is shown in Fig. 11. There is a jump in the rate allocation vector in going from the syndrome subchannels to information subchannels, as may be expected. Finally, it may be of interest that, for the Golay code with  $G = [P^T \ I_{12}]$ , we obtain  $\bar{R}_0(U_1^{23}; Y_1^{23}) = .4473$ , which is worse than the performance achieved by the dual code.

It is natural to ask at this point whether linear label maps can achieve normalized sum cutoff rates arbitrarily close to channel capacity in the limit as  $n$  goes to infinity. The next section shows that this is indeed possible.



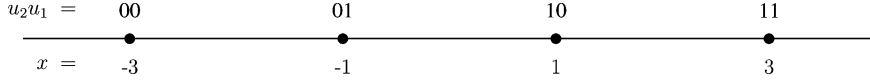


Fig. 12. Bit mapping for 4-PAM.

### V. PINSKER'S SCHEME

In this section, we show that the general framework of Section II admits Pinsker's method as a special case. This section will also help explain the relation of the present work to concatenated coding schemes in general.

In a theoretical study of coding complexity, Pinsker [10] combined the idea of sequential decoding and Elias' method of iterative coding [14] to prove that the cutoff rate of any DMC could be boosted to as close to channel capacity as desired. Pinsker's method is a form of serially concatenated coding where an inner block code is combined with an outer sequential decoder.

For simplicity, we describe Pinsker's method for the case of a BSC with a crossover probability  $0 \leq \epsilon \leq 1/2$ . Let  $R_0(\epsilon)$  and  $C(\epsilon)$  denote, respectively, the cutoff rate and capacity of a BSC with crossover probability  $\epsilon$ , and note that  $R_0(\epsilon) \rightarrow C(\epsilon)$  as  $\epsilon \rightarrow 0$ .

The encoding in Pinsker's scheme follows the method of Section IV-B. A systematic block code  $\mathcal{C}$  with rate  $R' = k/n$  and a generator matrix  $G = [P \ I_k]$  is chosen and a linear map  $F$  is formed as in (2). The decoder in Pinsker's scheme is a simplified form of the successive cancellation decoder where the decoders work completely independently, with no exchange of decisions. This is made possible by using a step-function-type rate allocation vector  $(R_1, \dots, R_n)$  such that  $R_1 = \dots = R_{n-k} = 0$  and  $R_{n-k+1} = \dots = R_n = R''$  where  $R''$  is to be specified. Thus,  $u_1^{n-k} \equiv 0$  and there is no need to decode the syndrome channels. The decoder at stage  $i$ ,  $n - k + 1 \leq i \leq n$ , computes the vector  $v_i^{n-k}$  as in (3), which directly reveals the syndrome vector  $s_1^{n-k}$  since  $u_1^{n-k} = 0$  by construction. Thus, each such decoder can independently implement the ML decoding rule using a syndrome decoding table. This creates  $k$  binary channels  $u_i \rightarrow \hat{u}_i$ ,  $n - k + 1 \leq i \leq n$ , where  $\hat{u}_i$  denotes the estimate of  $u_i$  obtained from the syndrome decoding table. By well-known coding theorems, there are codes  $\mathcal{C}$  for which the probability of error  $\Pr(\hat{u}_i \neq u_i)$  can be made arbitrarily small by using a sufficiently large  $n$ , provided that  $R' = k/n$  is kept bounded away from  $C(\epsilon)$ . (In fact,  $\Pr(\hat{u}_i \neq u_i)$  goes to zero exponentially in  $n$  for optimal codes.) Thus, the cutoff rate of each information subchannel, given by  $R_0(U_i, Y_1^n)$  under the independent decoding scheme described here, goes to 1.

To summarize, asymptotically as  $n \rightarrow \infty$ , we can choose  $R'$  arbitrarily close to  $C(\epsilon)$  and  $R''$  arbitrarily close to 1, resulting in an overall coding rate  $R'R''$  arbitrarily close to  $C(\epsilon)$ . This completes Pinsker's argument for improving the cutoff rate to channel capacity limit.

Clearly, Pinsker's argument is designed to give the proof of a theoretical point in the simplest possible form, with no attention to asymptotically insignificant details. In a nonasymptotic setting, the performance of the sequential decoders in Pinsker's scheme can be improved if the decoder at stage  $i$ ,  $n - k + 1 \leq i \leq n$ , calculates the soft-decision statistics

$$\Pr(u_i = b | y_1^n), \quad b = 0, 1$$

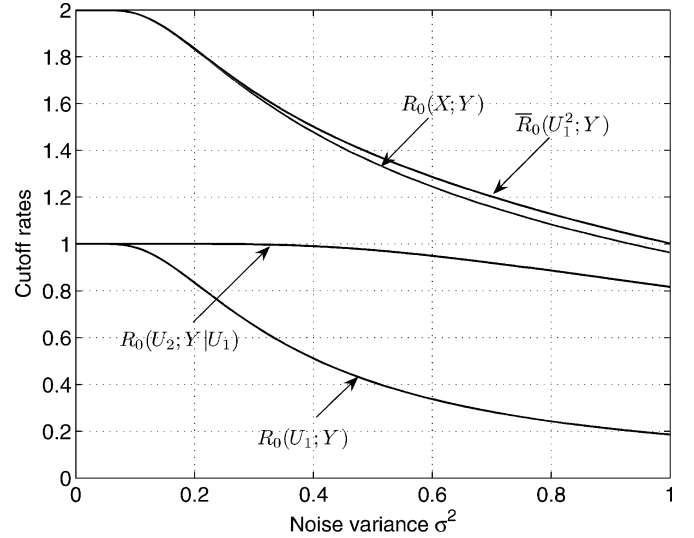


Fig. 13. Cutoff rates versus noise variance in 4-PAM.

instead of the hard decisions obtained from the syndrome decoder. The calculation of the soft-decision statistics may be performed by applying the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [20] to a trellis description of the block code  $\mathcal{C}$ .

In closing this section, we would like to mention Falconer's work [21], where another method is given for boosting the sum cutoff rate to near channel capacity. Falconer's method uses multiple inner sequential decoders and an outer Reed–Solomon code, and it does not fit into the framework considered in this paper. Neither Pinsker's method nor Falconer's appear to have had much impact on coding practice because of their complexity.

### VI. RELATION TO CODED MODULATION

The channel combining and splitting method presented in the preceding sections has common elements with well-known coded-modulation techniques, namely, Imai and Hirakawa's [13] multilevel coding scheme and Ungerboeck's [11], [12] set-partitioning idea. To illustrate this connection, we give the following example where the cutoff rate of a pulse-amplitude modulation (PAM) scheme is improved through channel splitting.

*Example 7 (4-PAM):* Consider a 4-PAM scheme over a memoryless additive Gaussian noise channel so that in each use of the channel a symbol  $x \in \{\pm 3, \pm 1\}$  is transmitted and  $y = x + n$  is received where  $n$  is Gaussian with mean zero and variance  $\sigma^2$ . In order to split this channel into two subchannels, the channel input  $x$  is relabeled by a pair of bits  $(u_2, u_1)$  as in Fig. 12 so that  $x = 4u_2 + 2u_1 - 3$ .

We consider an input ensemble so that  $U_1, U_2$  are independent and each takes the values 0 and 1 with probability  $1/2$ . The channel input ensemble  $X$  is then uniform over  $\{\pm 3, \pm 1\}$ . Fig. 13 shows the resulting cutoff rates  $R_0(X; Y)$ ,  $R_0(U_1; Y)$ ,

$R_0(U_2; Y|U_1)$ , and the normalized sum cutoff rate  $\bar{R}_0(U_1^2; Y)$ . We observe that this simple scheme improves the the cutoff rate.

Note that there is no channel combining in this example; the channel is split into two subchannels just by relabeling the input symbols, as in Massey's example. The assignment of bit strings to modulation symbols corresponds to Ungerboeck's set partitioning idea. The successive cancellation decoder first decodes  $U_1$ , the variable that is the less well protected of  $U_1$  and  $U_2$ . In fact, if one starts decoding starting with  $U_2$ , the cutoff rate deteriorates:  $R_0(U_2; Y) + R_0(U_1; Y|U_2) < R_0(X; Y)$ . To attain further cutoff rate gains, one may try combining multiple copies of the channel prior to splitting.

The preceding example illustrates the close connection between the present work and the coded-modulation subject. Indeed, this connection is exploited in the paper by Wachsmann *et al.* [22], where the authors give practical design rules for multilevel codes using various figures of merit, including the sum cutoff rate. Persson [23] is credited in [22] for noticing that the sum cutoff rate of a modulation scheme under multilevel coding may exceed the ordinary cutoff rate. However, Massey's work [1] is clearly an earlier example of such "anomalous" cutoff rate phenomenon in the context of modulation.

## VII. RELIABILITY-COMPLEXITY TRADEOFF

In this section, we show that the channel combining and splitting method discussed above can also improve the reliability-complexity tradeoff in channel coding. For background on the topic of reliability-complexity tradeoff in ML decoding, we refer to [24] and [9, Sec. 6.6].

The formulation in this section is based on the following functions. For  $\rho \geq 0$  and  $(X, Y) \sim P(x, y)$ , define

$$E_{0,\rho}(X; Y) = -\log \sum_y \left[ \sum_x P(x) P(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

and for  $(X, Z, Y) \sim P(x, z, y)$ , define

$$E_{0,\rho}(X; Y|Z) = -\log \sum_{z,y} P(z) \left[ \sum_x P(x|z) P(y|x, z)^{\frac{1}{1+\rho}} \right]^{1+\rho}.$$

These functions appear in random-coding exponents for block coding on both single- and multiuser channels [8, Ch. 5], [25], [2]. They equal the  $R_0$  function when  $\rho = 1$ . Also, for any fixed ensemble  $(X, Z, Y)$ ,  $E_{0,\rho}(X; Y|Z)/\rho$  is a monotonically non-increasing function of  $\rho > 0$ , and it tends to the mutual information  $I(X; Y|Z)$  as  $\rho \rightarrow 0$ . Given this limit relation, it is not surprising that the function  $E_{0,\rho}(X; Y|Z)$  possesses many properties of the mutual information. However, unlike mutual information, the function  $E_{0,\rho}$  is not additive; i.e., there are examples for which  $E_{0,\rho}(X; Y) + E_{0,\rho}(Z; Y|X) \neq E_{0,\rho}(XZ; Y)$ . In this section, we show by an example that we may take advantage of the lack of additivity of  $E_{0,\rho}$  to improve the reliability-complexity tradeoff in ML decoding. We begin by defining the reliability-complexity exponent.

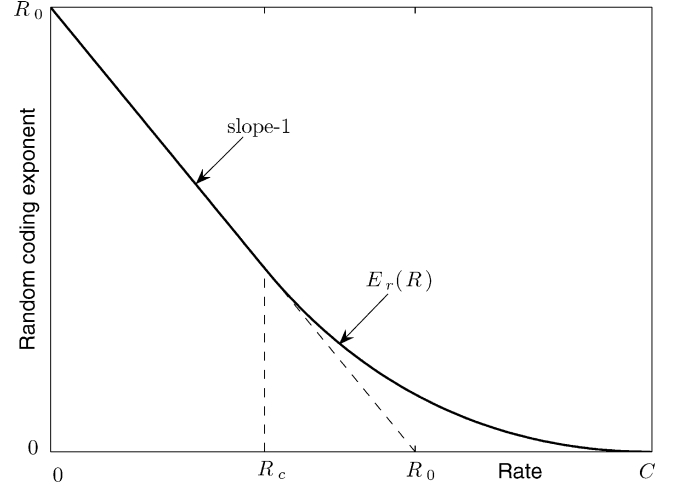


Fig. 14. Typical random-coding exponent.

### A. Reliability-Complexity Exponent

Consider a DMC  $V : \mathcal{X} \rightarrow \mathcal{Y}$  and let  $Q$  be a probability distribution on  $\mathcal{X}$ . For  $R \geq 0$ , define

$$E_r(R, Q, V) = \max_{0 \leq \rho \leq 1} [E_{0,\rho}(X; Y) - \rho R]$$

where  $(X, Y) \sim Q(x)V(y|x)$ . Also define

$$E_r(R, V) = \max_Q E_r(R, Q, V).$$

The function  $E_r(R, V)$  is called the *random-coding exponent* [8, p. 143]. Fig. 14 shows a typical random-coding exponent. The exponent is positive for all rates between zero and channel capacity  $C$ . The slope of the exponent equals  $-1$  for a range of rates  $0 \leq R \leq R_{cr}$  where  $R_{cr}$  is called the critical rate. The slope increases monotonically for  $R_{cr} \leq R \leq C$  and approaches 0 as  $R$  approaches  $C$ . The vertical axis intercept of the random-coding exponent is given by the cutoff rate  $R_0(V)$ .

Gallager [26], [8, Theorem 5.6.2] shows that  $\bar{P}_e$ , the probability of ML decoding error averaged over an  $(N, 2^{NR}, Q)$  block code ensemble, is upper-bounded by

$$\bar{P}_e \leq 2^{-NE_r(R, Q, V)}.$$

An  $(N, 2^{NR}, Q)$  block code ensemble consists of the set of all block codes of length  $N$  and  $M = \lceil 2^{NR} \rceil$  codewords where we assign the probability  $\prod_{m=1}^M \prod_{n=1}^N Q(x_{m,n})$  to the code  $\mathcal{C} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ ,  $\mathbf{x}_m = (x_{m,1}, \dots, x_{m,N})$ ,  $1 \leq m \leq M$ . Gallager [27] also shows that the exponent  $E_r(R, Q, V)$  is tight in the sense that  $-(1/N) \log \bar{P}_e$  goes to  $E_r(R, Q, V)$  as  $N$  increases. We write  $\bar{P}_e \cong 2^{-NE_r(R, Q, V)}$  to express this type of asymptotic equality.

The complexity of ML decoding of a code from ensemble  $(N, 2^{NR}, Q)$  is given by  $\chi \cong 2^{NR}$ . So, the reliability-complexity tradeoff under ML decoding of codes from this ensemble may be expressed as  $\bar{P}_e \cong \chi^{-E_r(R, Q, V)/R}$ . Thus, the parameter  $E_r(R, Q, V)/R$  emerges as the *reliability-complexity exponent* for ML decoding of a typical code from the ensemble  $(N, 2^{NR}, Q)$ . Maximizing  $E_r(R, Q, V)/R$  over  $Q$ , we obtain  $E_r(R, V)/R$  as the best attainable reliability-complexity exponent at rate  $R$ .

### B. Improvement of the Reliability–Complexity Exponent

We now show that under certain conditions, the reliability–complexity tradeoff of a given DMC may be improved. Consider a DMC  $V : \mathcal{X} \rightarrow Y$ , and fix a rate  $0 < R < C(V)$ . Let  $Q^*$  and  $\rho^*$  be such that  $E_r(R, V) = E_{0,\rho^*}(Q^*, V) - \rho^*R$ .

Suppose that two copies of  $V$  can be combined in accordance with the general method in Fig. 4 to yield a channel  $W : U_1^2 \rightarrow Y_1^2$  such that

$$E_{0,\rho^*}(U_1; Y_1^2) + E_{0,\rho^*}(U_2; Y_1^2|U_1) > 2E_{0,\rho^*}(Q^*; V) \quad (4)$$

for some probability assignment  $U_1^2 \sim Q_1(u_1)Q_2(u_2)$ . Let  $R_1, R_2$  be defined so that  $R_1 + R_2 = 2R$  and

$$\frac{E_{0,\rho^*}(U_1; Y_1^2)}{R_1} = \frac{E_{0,\rho^*}(U_2; Y_1^2|U_1)}{R_2}. \quad (5)$$

Due to (4), it is easy to show that the ratios in (5) are greater than  $E_{0,\rho^*}(Q^*, V)/R$ .

The significance of these ratios become evident when we consider a successive cancellation scheme as in Fig. 5, where the two inputs of  $W$  are encoded at rates  $R_1$  and  $R_2$  using the ensemble  $U_1^2 \sim Q_1(u_1)Q_2(u_2)$ . Then, the reliability–complexity exponents for decoders 1 and 2 are given, respectively, by

$$\frac{\max_{0 \leq \rho \leq 1} [E_{0,\rho}(U_1; Y_1^2) - \rho R_1]}{R_1}$$

and

$$\frac{\max_{0 \leq \rho \leq 1} [E_{0,\rho}(U_2; Y_1^2|U_1) - \rho R_2]}{R_2},$$

both of which are greater than  $E_{0,\rho^*}(Q^*, V)/R$ . Thus, whenever (4) is satisfied, using ML decoders in a successive cancellation configuration offers a better reliability–complexity tradeoff compared to ordinary ML decoding.

The preceding argument clearly generalizes to the case where an arbitrary number of copies of a given DMC  $V$  are combined. In the next subsection, we revisit Massey's example to illustrate the improvement of the reliability–complexity exponent by channel splitting.

### C. Reliability Versus Complexity in Massey's Example

We consider the two coding alternatives for a QEC as shown in Fig. 2. For the first alternative, the reliability–complexity exponent is  $E_r(R, \text{QEC})/R$ . For the second alternative, the exponent for each decoder is  $E_r(R/2, \text{BEC})/(R/2)$ . Thus, we will have an improvement in the reliability–complexity tradeoff if

$$2E_r(R/2, \text{BEC}) > E_r(R, \text{QEC}). \quad (6)$$

Humblet [28] gives the random-coding exponent for the  $M$ -ary erasure channel (MEC) as follows:

$$E_r(R, \text{MEC}) = \begin{cases} D(1 - \frac{R}{\log M} \|\epsilon), & R_{\text{cr}} \leq R \leq C \\ R_0 - R, & 0 \leq R \leq R_{\text{cr}} \end{cases} \quad (7)$$

where  $D(\delta \|\epsilon) = \delta \log(\delta/\epsilon) + (1 - \delta) \log[(1 - \delta)/(1 - \epsilon)]$ . The other parameters in (7) are given by  $C = (1 - \epsilon) \log M$ ,  $R_{\text{cr}} = C/[1 + (M - 1)\epsilon]$ , and  $R_0 = \log M - \log[1 + (M - 1)\epsilon]$ ; these are the capacity, critical rate, and cutoff rate, respectively, of the MEC. Evaluating (7) for  $M = 4$  and  $M = 2$ , we obtain the random-coding exponents for the QEC and BEC, respectively. Fig. 15 is a plot of  $E_r(R, \text{QEC})$  and  $2E_r(R/2, \text{BEC})$

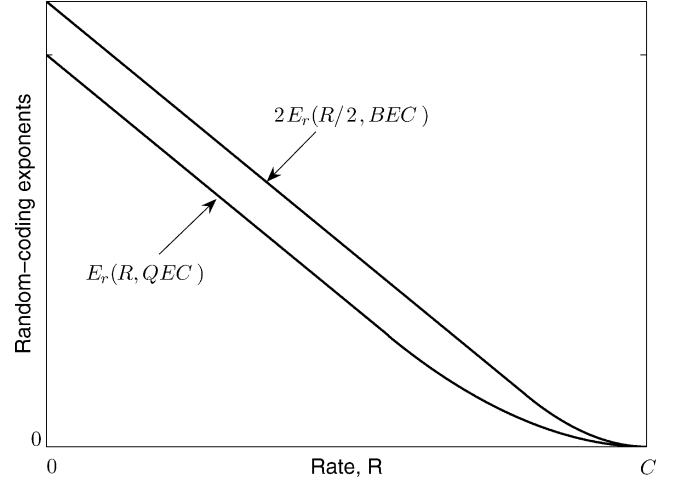


Fig. 15. Random-coding exponents for QEC and BEC.

for  $\epsilon = 0.25$ . It is seen from the figure and can be verified directly that (6) is satisfied for all  $0 \leq R < C(\text{QEC})$ . In fact, for rates  $R_{\text{cr}}(\text{QEC}) \leq R \leq C(\text{QEC})$ , we have  $E_r(R/2, \text{BEC}) = E_r(R, \text{QEC})$ ; so, the reliability–complexity exponent is doubled by splitting. Thus, for the same order of decoder complexity, splitting the QEC into two BECs offers significantly higher reliability, especially at high rates.

### D. Improving Other Exponents

We should mention that the channel combining and splitting method can also be used to improve the reliability–complexity tradeoff in convolutional coding, for which ML decoding (Viterbi decoding) is a widely used technique. The only difference then is that one needs to consider the Yudkin–Viterbi exponent [29], [30], which is the ML decoding exponent for (time-varying) convolutional codes.

We also mention that channel splitting may improve the reliability–complexity tradeoff as computed by the *expurgated exponent* (see [8, p. 153] for definition), which improves the random-coding exponent at rates below the critical rate  $R_{\text{cr}}$ .

## VIII. CONCLUDING REMARKS

We have discussed a method for improving the sum cutoff rate of a given DMC. Although the method has been presented for some binary-input channels, it is readily applicable to a wider class of channels. The method is based on combining a number of independent copies of a given channel into a number of correlated parallel channels through input relabeling. Encoding of the subchannels is carried out independently but decoding is done using a successive cancellation-type decoder. Some elements of this method may be traced back to Elias' iterative coding idea, some to concatenated coding, coded modulation, and multilevel decoding. These connections have also been discussed.

Some important questions have not been addressed in this study. We have illustrated through several examples the effectiveness of the method in improving the sum cutoff rate. However, we have not addressed theoretical questions of why the

cutoff rate is improved. More fundamental questions of this nature are left for future study.

Our main aim has been to explore the existence of practical schemes that boost the sum cutoff rate to near channel capacity. This goal remains far from being achieved. The main obstacle in this respect is the complexity of successive cancellation decoding which grows exponentially with  $n$ , the number of channels combined. However, it is conceivable that decoding schemes that approximate successive cancellation decoding exist that overcome this difficulty. This is another issue that may be studied further.

## REFERENCES

- [1] J. L. Massey, "Capacity, cutoff rate, and coding for a direct-detection optical channel," *IEEE Trans. Commun.*, vol. COM-29, no. 11, pp. 1615–1621, Nov. 1981.
- [2] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 2, pp. 124–142, Mar. 1985.
- [3] J. M. Wozencraft and R. S. Kennedy, "Modulation and demodulation for probabilistic coding," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 4, pp. 291–297, Jul. 1966.
- [4] J. L. Massey, "Coding and modulation in digital communication," in *Proc. Int. Zurich Seminar on Digital Communication*, Zurich, Switzerland, 1974, pp. E2(1)–E2(24).
- [5] J. M. Wozencraft and B. Reiffen, *Sequential Decoding*. Cambridge, MA: MIT Press, 1961.
- [6] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 2, pp. 64–74, Apr. 1963.
- [7] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.
- [8] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [9] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
- [10] M. S. Pinsker, "On the complexity of decoding," *Probl. Pered. Inform.*, vol. 1, no. 1, pp. 113–116, 1965.
- [11] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets, Part I: Introduction," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 5–11, Feb. 1987.
- [12] ———, "Trellis-coded modulation with redundant signal sets, Part II: State of the art," *IEEE Commun. Mag.*, vol. 25, no. 2, pp. 12–21, Feb. 1987.
- [13] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
- [14] P. Elias, "Error-free coding," *IRE Trans. Inf. Theory*, vol. PGIT-4, no. 2, pp. 29–37, Sep. 1954.
- [15] G. D. Forney Jr., *Concatenated Codes*. Cambridge, MA: MIT Press, 1966.
- [16] V. V. Zyablov, "An estimate of complexity of constructing binary linear cascade codes," *Probl. Inf. Transm.*, vol. 7, no. 1, pp. 3–10, 1971.
- [17] E. L. Blokh and V. V. Zyablov, *Linear Concatenated Codes* (in Russian). Moscow, U.S.S.R.: Nauka, 1982.
- [18] E. W. Weisstein *et al.*, Golay Code. MathWorld—A Wolfram Web Resource. [Online]. Available: <http://mathworld.wolfram.com/Golay-Code.html>
- [19] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.
- [20] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.
- [21] D. D. Falconer, "A hybrid coding scheme for discrete memoryless channels," *Bell Syst. Tech. J.*, pp. 691–728, Mar. 1969.
- [22] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
- [23] J. Persson, "Multilevel Coding Based on Convolutional Codes," Ph.D. dissertation, Lund University, Lund, Sweden, 1996.
- [24] G. D. Forney Jr., "1995 Shannon Lecture—Performance and complexity," *IEEE Inf. Theory Soc. Newslett.*, vol. 46, pp. 3–4, Mar. 1996.
- [25] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.
- [26] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.
- [27] ———, "The random coding bound is tight for the average code (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 244–246, Mar. 1973.
- [28] P. Humblet, "Error Exponents for Direct Detection Optical Channel," MIT, Lab. Information and Decision Systems, Rep. LIDS-P-1337, 1983.
- [29] H. L. Yudkin, "Channel State Testing in Information Decoding," Ph.D. dissertation, MIT, Dept. Elect. Eng., Cambridge, MA, 1964.
- [30] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 260–269, Apr. 1967.